Gene, 124 (1993) 45–55
© 1993 Elsevier Science Publishers B.V. All rights reserved. 0378-1119/93/$06.00

P.D. 1993
P. 45-55 M

GENE 06930

XP-001084201

Cloning and analysis of *Candida cylindracea* lipase sequences

(Yeast; gene structure; fungal extracellular lipase; homology; evolution)

Marina Lotti[a], Rita Grandori[a], Fabrizia Fusetti[a], Sonia Longhi[a], Stefania Brocca[a], Anna Tramontano[b] and Lilia Alberghina[a]

[a]*Dipartimento di Fisiologia e Biochimica Generali, Università degli Studi di Milano, 20133 Milano, Italy; and* [b] *IRBM P. Angeletti, Via Pontina Km 30.600, 0045 Pomezia, Roma, Italy. Tel. (39-6) 91093207*

Received by J.K.C. Knowles: 21 May 1992; Revised/Accepted: 17 September/28 September 1992; Received at publishers: 10 November 1992

SUMMARY

Lipases (Lip) hydrolyze triglycerides into fatty acids and glycerol. Lip produced by the yeast *Candida cylindracea* are encoded by multiple genomic sequences. We report the molecular cloning and characterization of three genes from this family. They encode putative mature 57-kDa proteins of 534 amino acids (aa). To date, five Lip-encoding genomic sequences from *C. cylindracea* have been characterized in our laboratory. The five deduced aa sequences share an overall homology of 80%. These sequences have been aligned with each other and with those of homologous enzymes, the Lip from the mould *Geotrichum candidum* and the acetylcholinesterase from *Torpedo californica*, whose three-dimensional structures have been solved by X-ray analysis. The *C. cylindracea* Lip appear to have a structural organization similar to that described for both enzymes.

INTRODUCTION

Lipases (Lip; triacylglycerol acyl hydrolases) are key enzymes in fat metabolism that catalyze the breakdown of triacylglycerols to free fatty acids and glycerol. They constitute an ubiquitous group of enzymes able to catalyze a number of different reactions, many of them of industrial interest (reviewed in Harwood, 1989). Lip hydrolyze ester bonds of water-insoluble substrates at the interface between the organic phase containing the emulsified substrate and the aqueous phase in which the enzyme is soluble. Lip activity is largely increased at the lipid-water interface (Brockman, 1984).

The biochemical and molecular characterization of a number of Lip obtained from different sources has pointed to a marked heterogeneity of this class of enzymes with regard to specificity, aa sequence and catalytic properties (Antonian, 1988). Based initially on the inhibition of enzymatic activity by chemical modification, Lip have been classified in the Ser hydrolases group (Chapus et al., 1976). The serine of the active site has been shown to be enclosed in a highly conserved Gly-Xaa-Ser-Xaa-Gly motif. This consensus domain represents the only feature shared by all determined lipase sequences (Antonian, 1988).

Recently, a very significant contribution to the understanding of the Lip structure and function relationships arose from the elucidation of the three-dimensional structure of Lip from the human pancreas (Winkler et al., 1990), *Rhizomucor miehei* (Brady et al., 1990) and *Geotrichum candidum* (Gcl; Schrag et al., 1991). The three enzymes differ in their sequences (Boel et al., 1988;

Correspondence to: Dr. L. Alberghina, Dipartimento di Fisiologia e Biochimica Generali, Via Celoria 26, 20133 Milano, Italy. Tel. (39-2) 70644801; Fax (39-2) 2361070.

Abbreviations: aa, amino acid(s); AChE, acetylcholinesterase(s); bp, base pair(s); *C., Candida*; Ccl, lipase from *C. cylindracea*; ChE, cholinesterase(s); *G., Geotrichum*; Gcl, lipase from *G. candidum*; kb, kilobase(s) or 1000 bp; Lip, lipase(s); *LIP*, Lip-encoding gene(s); nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; pI, isoelectric point; *T., Torpedo*; TAChE, acetylcholinesterase from *T. californica*.

Shimada et al., 1989; 1990; Winkler et al., 1990) and in several structural features. Nevertheless, significant similarities have been observed in the topology of the mixed β-sheet conserved in the protein hydrophobic core and in the configuration of the Ser-His-Asp/Glu catalytic triad, functionally equivalent to that described for Ser proteases. The active site is buried in a deep hydrophobic cavity covered by amphipathic helical elements named by the authors 'lid' or 'flap'. Interfacial activation appears to imply a conformational change of the 'lid', making the active site accessible to the substrate. This suggested mechanism finds support in the x-ray studies performed on a complex of *Rhizomucor* lipase with a substrate analogue, where the flap appeared rotated by about 167° (Brzozowski et al., 1991).

The extracellular Lip produced by the asporogenic yeast *C. cylindracea* (Ccl) hydrolyzes triglycerides without specificity both in the attacked position of the glycerol molecule and in the nature of the fatty acid released. This enzyme has therefore a relaxed specificity in comparison with other Lip. This property makes it useful for industrial applications. In previous reports from our laboratory, it has been shown that multiple Lip-related sequences are present in the *C. cylindracea* genome (Alberghina et al., 1991), and the sequences of two of them (*LIP1* and *LIP2*) have been determined (Longhi et al., 1992). In this paper we report the cloning and sequencing of three other Lip genes and an analysis of the sequences of the *C. cylindracea LIP* multigene family, compared with each other and with those of related enzymes. On the basis of these results, we propose a model for the structural organization of the *C. cylindracea* Lip.

RESULTS AND DISCUSSION

(a) Cloning of three genomic *LIP* sequences

C. cylindracea genomic DNA was digested to completion with restriction endonucleases and subjected to Southern blot analysis using fragments derived from *LIP1* (Longhi et al., 1992) as probes (fig. 1A). Duplicate blots were hybridized to probes specific for either the 5'-end or the 3'-end of the gene. Multiple hybridization signals were observed (Fig. 1B). In the pattern obtained with *Sac*I-cleaved DNA, four bands were positive to both probes (not shown). The two *Sac*I bands of about 4.2 and 5.1 kb (Fig. 1B) are likely to correspond to the two genomic clones previously isolated in our laboratory (Longhi et al., 1992). Moreover, the complexity of the hybridization pattern indicated the presence of other closely related homologous sequences. Therefore, total yeast DNA was cleaved with *Sac*I, size selected for fragment lengths >7 kb and used to construct another genomic library. Screening of ca. 7000 recombinants led to the isolation of three positive clones carrying 10 kb *Sac*I-*Sac*I inserts. One of them was further characterized by a fine hybridization analysis using five probes representative of the whole *LIP1* sequence, as indicated in Fig. 1A. This clone was found to contain three sequences apparently oriented head to tail, hybridizing to all five probes and showing distinctive restriction patterns (Fig. 1C).

(b) Nucleotide sequence of *LIP* and deduced aa sequence

LIP-homologous sequences (*LIP3*, *LIP4* and *LIP5*) were separately subcloned into pGEM3Z (Promega, Madison, WI) and their complete nt sequences were determined on both strands (Fig. 2). The nt sequences obtained were translated into the corresponding aa sequences taking into account the unusual usage of the codon CTG—Ser instead of Leu—occurring in this *Candida* strain (Kawaguchi et al., 1989).

The *LIP3* sequence contains a unique ORF of 1647 nt terminating with a TAG stop codon (Fig. 2A). The deduced aa sequence corresponds to a protein of 549 aa with an N-terminal stretch of 15 hydrophobic aa, which might encode a signal peptide. The cleavage of the preprotein by the signal peptidase is predicted to occur after the tripeptide ValAlaAla (von Heijne, 1986). *LIP3* would therefore produce a mature protein of 534 aa and M_r 57 291. Three potential *N*-glycosylation sites (Asn-Xaa-Ser/Thr; Hubbard and Ivatt, 1981) have been identified at aa positions 291, 314 and 351. Five Cys residues are present at aa positions 60, 97, 217, 268 and 277. The predicted pI (isoelectric point) of the Lip3 protein is 5.1.

The *LIP4* sequence spans an ORF of 1647 nt closed by a TAG codon, corresponding to a preprotein of 549 aa (Fig. 2B). Upon cleavage of the signal peptide this would result in a mature protein of 534 aa and M_r 57 051. The putative Lip4 contains a unique site for *N*-glycosylation (Asn^{351}), and six Cys at positions 60, 97, 217, 268, 277 and 328. Its calculated pI is 5.7.

The *LIP5* ORF is 1647 nt long and is terminated by a TAG stop codon, corresponding to a mature protein of 534 aa and M_r 56.957, preceded by a leader peptide of 15 aa. Lip5 contains two putative *N*-glycosylation sites (Asn^{314} and Asn^{351}) and five Cys at position 60, 97, 217, 268 and 277. Its calculated pI is 5.5.

(c) Comparison of the deduced aa sequences

LIP3, *LIP4* and *LIP5* encode for putative proteins of 534 aa highly related in sequence to each other and to the Lip1 and Lip2 from the same organism (Longhi et al., 1992). All five *LIP* sequences will be considered in the following analyses. Residues conserved in all five sequences make up 66% if identity is considered and 84%

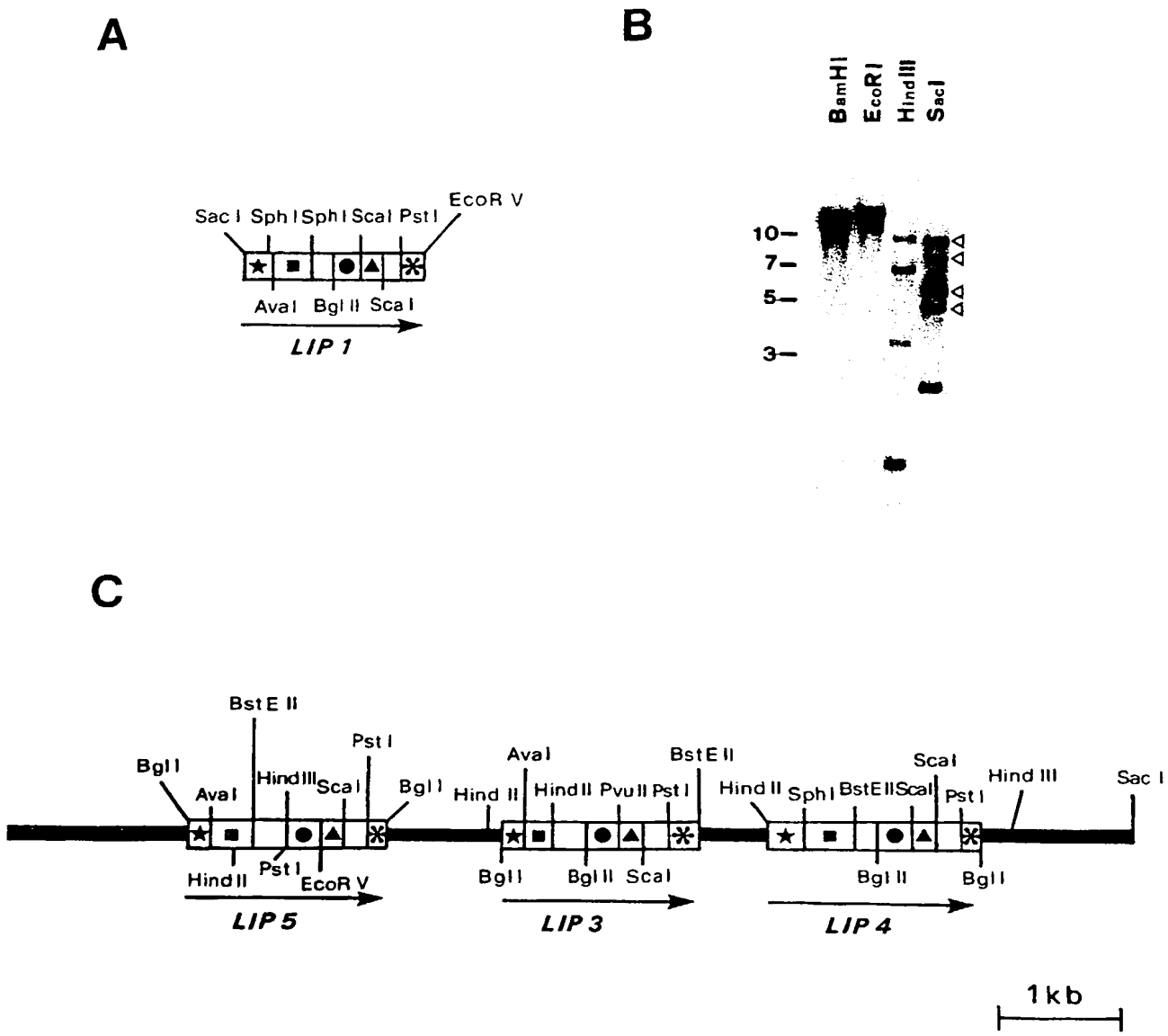


Fig. 1. *C. cylindracea* genome analysis and cloning of *LIP*. (A) Location of the probes (ca. 200 bp) in *LIP1*. DNA fragments were purified from agarose gels and labeled by random priming with [^{32}P]dCTP. Probes are marked by different symbols. (Panel B) Southern blots of genomic DNA digested with restriction endonucleases and hybridized to the *Sac*I-*Ava*I probe specific for the 5′-end of *LIP*. A similar hybridization experiment was performed with the *Pst*I-*Eco*RV probe specific for the 3′-end (not shown). *Sac*I bands positive to both probes are indicated by open triangles. Hybridization was carried out at 45°C. (C) Location of *LIP* homologous sequences within the 10-kb *Sac*I-*Sac*I fragment as deduced by Southern analysis with the probes described in A. Hybridization conditions were as above. DNA regions positive to the same probe are indicated by the same symbol. Only restriction sites used for this analysis are reported. Coding sequences and their orientation are represented by arrows.

if aa are grouped on the basis of their physicochemical similarity as specified in the SIMPLIFY program of the GCG sequence analysis package (Devereux et al., 1984). The Lip consensus sequence Gly-Xaa-Ser-Xaa-Gly is conserved in all sequences (Fig. 2). A common feature of the putative Lip is the presence of hydrophobic signal peptides, whose cleavage would generate mature proteins with Ala as the first aa. These observations are in agreement with the results obtained by sequencing the N terminus of the mature protein (Kawaguchi et al., 1989). The M_r predicted for the Lip are in agreement with those experimentally estimated by gel electrophoresis of purified Ccl commercial preparations (Veeraragavan and Gibbs, 1989). Potential sites of glycosylation, one of which (Asn^{351}) is conserved in all five proteins, have been identified. Biochemical analyses have shown that Ccl contains 4% sugars, mainly mannose and xylose (Tomizuka et al., 1966). Four Cys residues are conserved, which might participate in the formation of disulfide bridges.

Analysis of 5′ regions showed the presence of TATAA

A

```
                                                    GTCGACAACAATGCGGCGCTGATCGCGGACTACCGCACACCGCTCCGAGTATAAGCAGAAGCATTCTCACCTGCTCGCTCCCC

ATG AAG CTC GCT CTT GCG CTC CTG CTC ATT GCC TCG GTG GCT GCC GCC CCC ACC GCC AAG CTC GCC AAC GGC GAC   75
Met Lys Leu Ala Leu Ala Leu Ser Leu Ile Ala Ser Val Ala Ala Ala Pro Thr Ala Lys Leu Ala Asn Gly Asp   10

ACC ATC ACC GGT CTC AAC GCC ATC ATC AAC GAG GCG TTC CTC GGC ATT CCC TTT GCC GAG CCG CCG GTG GGC AAC   15
Thr Ile Thr Gly Leu Asn Ala Ile Ile Asn Glu Ala Phe Leu Gly Ile Pro Phe Ala Glu Pro Pro Val Gly Asn   35

CTC CGC TTC AAG GAC CCT GTG CCG TAC TCT GGC TCG CTC AAC GGC CAG AAG TTT ACT CTG TAC GGC CCG CTG TGC   22
Leu Arg Phe Lys Asp Pro Val Pro Tyr Ser Gly Ser Leu Asn Gly Gln Lys Phe Thr Ser Tyr Gly Pro Ser Cys   60

ATG CAG CAG AAC CCC GAG GGC ACG TTT GAA GAG AAC CTT GGC AAG ACG GCA CTC GAC TTG GTG ATG CAG TCC AAG   30
Met Gln Gln Asn Pro Glu Gly Thr Phe Glu Glu Asn Leu Gly Lys Thr Ala Leu Asp Leu Val Met Gln Ser Lys   85

GTG TTC CAG GCG GTG CTT CCC CAG AGT GAG GAC TGC CTC ACC ATC AAC GTG GTG CGG CCG CCG GGC ACC AAG GCG   37
Val Phe Gln Ala Val Leu Pro Gln Ser Glu Asp Cys Leu Thr Ile Asn Val Val Arg Pro Pro Gly Thr Lys Ala   11

GGC GCC AAC CTC CCG GTC ATG CTC TGG ATC TTT GGC GGT GGG TTT GAG ATC GGC AGC CCC ACC ATC TTC CCT CCC   45
Gly Ala Asn Leu Pro Val Met Leu Trp Ile Phe Gly Gly Gly Phe Glu Ile Gly Ser Pro Thr Ile Phe Pro Pro   13

GCC CAG ATG GTC ACC AAG AGT GTG CTC ATG GGC AAG CCC ATC ATC CAC GTG GCC GTC AAC TAC CGT GTT GCC TCG   52
Ala Gln Met Val Thr Lys Ser Val Leu Met Gly Lys Pro Ile Ile His Val Ala Val Asn Tyr Arg Val Ala Ser   16

TGG GGG TTC TTG GCT GGT GAT GAC ATC AAG GCC GAG GGC AGC GGG AAC GCC GGC TTG AAG GAC CAG CGT TTG GGC   60
Trp Gly Phe Leu Ala Gly Asp Asp Ile Lys Ala Glu Gly Ser Gly Asn Ala Gly Leu Lys Asp Gln Arg Leu Gly   18

ATG CAG TGG GTG GCA GAC AAC ATT GCC GGG TTC GGC GGC GAC CCG AGC AAG GTG ACG ATC TTT GGC GAG CTG GCG   67
Met Gln Trp Val Ala Asp Asn Ile Ala Gly Phe Gly Gly Asp Pro Ser Lys Val Thr Ile Phe Gly Glu Ser Ala   21
                                                                                  *   *   *   *
GGC AGC ATG TCC GTG TTG TGC CAC CTC ATC TGG AAC GAC GGC GAC AAC ACG TAC AAG GGC AAG CCG TTG TTC CGC   75
Gly Ser Met Ser Val Leu Cys His Leu Ile Trp Asn Asp Gly Asp Asn Thr Tyr Lys Gly Lys Pro Leu Phe Arg   23
 *
GCG GGC ATC ATG CAG CTG GGA GCC ATG GTG CCG CTG GAC CCG GTG GAC GGC ACG TAC GGC AAC GAG ATC TAC GAC   82
Ala Gly Ile Met Gln Ser Gly Ala Met Val Pro Ser Asp Pro Val Asp Gly Thr Tyr Gly Asn Glu Ile Tyr Asp   26

CTC TTT GTC TCG AGT GCT GGC TGT GGC AGC GCC AGC GAC AAG CTC GCG TGC TTG CGC AGT GCG CTG AGC GAC ACC   90
Leu Phe Val Ser Ser Ala Gly Cys Gly Ser Ala Ser Asp Lys Leu Ala Cys Leu Arg Ser Ala Ser Ser Asp Thr   28

TTG CTC GAT GCC ACC AAC AAC ACT CCT GGG TTC TTG GCG TAC TCC TCG TTG CGG TTG CTG TAC CTT CCC CGG CCC   97
Leu Leu Asp Ala Thr Asn Asn Thr Pro Gly Phe Leu Ala Tyr Ser Ser Leu Arg Leu Ser Tyr Leu Pro Arg Pro   31

GAC GGC AAG AAC ATC ACC GAT GAC ATG TAC AAG TTG GTG CGC GAC GGC AAG TAT GCA AGC GTT CCC GTG ATC ATT   10
Asp Gly Lys Asn Ile Thr Asp Asp Met Tyr Lys Leu Val Arg Asp Gly Lys Tyr Ala Ser Val Pro Val Ile Ile   33

GGC GAC CAG AAC GAC GAG GGC ACC ATC TTT GGT CTC CTG CTG TTG AAC GTG ACC ACG AAT GCT CAG GCC CGT GCT   11
Gly Asp Gln Asn Asp Glu Gly Thr Ile Phe Gly Leu Ser Ser Leu Asn Val Thr Thr Asn Ala Gln Ala Arg Ala   36

TAC TTC AAG CAG CTG TTC ATC CAC GCC AGC GAC GCG GAG ATC GAC ACC TTG ATG GCG GCG TAC CCC CAG GAC ATC   12
Tyr Phe Lys Gln Ser Phe Ile His Ala Ser Asp Ala Glu Ile Asp Thr Leu Met Ala Ala Tyr Pro Gln Asp Ile   38

ACC CAG GGT CTG CCG TTC GAC ACC GGC ATC TTC AAC GCA ATC ACC CCG CAG TTC AAG AGA ATC CTG GCG GTG CTC   12
Thr Gln Gly Ser Pro Phe Asp Thr Gly Ile Phe Asn Ala Ile Thr Pro Gln Phe Lys Arg Ile Ser Ala Val Leu   41

GGC GAC CTT GCA TTC ATC CAC GCC CGC CGC TAC TTC CTC AAC CAC TTC CAG GGC GGC ACC AAG TAC TCG TTC CTC   13
Gly Asp Leu Ala Phe Ile His Ala Arg Arg Tyr Phe Leu Asn His Phe Gln Gly Gly Thr Lys Tyr Ser Phe Leu   43

CTG AAG CAG CTC CTG GGG TTG CCA ATC ATG GGC ACC TTC CAT GCC AAC GAC ATT GTG TGG CAG GAC TAC TTG TTG   14
Ser Lys Gln Leu Ser Gly Leu Pro Ile Met Gly Thr Phe His Ala Asn Asp Ile Val Trp Gln Asp Tyr Leu Leu   46

GGA AGC GGC AGC GTC ATC TAC AAC AAC GCG TTT ATC GCG TTC GCC ACC GAC TTG GAC CCC AAC ACC GCG GGG TTG   15
Gly Ser Gly Ser Val Ile Tyr Asn Asn Ala Phe Ile Ala Phe Ala Thr Asp Leu Asp Pro Asn Thr Ala Gly Leu   48

TTG GTG AAC TGG CCC AAG TAC ACC AGC AGC CTG CAG CTG GGC AAC AAC TTG ATG ATG ATC AAC GCC TTG GGC TTG   15
Leu Val Asn Trp Pro Lys Tyr Thr Ser Ser Ser Gln Ser Gly Asn Asn Leu Met Met Ile Asn Ala Leu Gly Leu   51

TAC ACC GGC AAG GAC AAC TTC CGC ACC GCT GGC TAC GAC GCG TTG ATG ACC AAC CCG CTG CTG TTC TTT GTG TAG   16
Tyr Thr Gly Lys Asp Asn Phe Arg Thr Ala Gly Tyr Asp Ala Leu Met Thr Asn Pro Ser Ser Phe Phe Val ooo   53

TTGTGTATGTGCCAGTATGGATATGTGTGATTCGGCTCCCCAAAACCCTGTATCCATCCATCACGCATCTTCAGCAAACGGTAACTTGGCGTATGCCGA

CAACCAGTGGACGTTGCTCCTGG
```

B

```
                                                    GTTAACAACAACTAGCTGATCGGCGGCGGGGGGCTATGCGCCGGCCTAAAATGTTCCGCTTATAACCGCGGACTCTCCACTCGCTCCAGCTCACTCCCC

ATG AAG CTC GCT CTT GTA CTC TCG CTC ATT GTC TCG GTG GCC GCG GCC CCC ACT GCC ACG CTC GCC AAC GGC GAC   75
Met Lys Leu Ala Leu Val Leu Ser Leu Ile Val Ser Val Ala Ala Ala Pro Thr Ala Thr Leu Ala Asn Gly Asp   10

ACC ATC ACC GGT CTC AAC GCC ATC ATC AAC GAG GCG TTC CTC GGT ATT CCC TTT GCT CAG CCG CCG GTG GGC AAC   150
Thr Ile Thr Gly Leu Asn Ala Ile Ile Asn Glu Ala Phe Leu Gly Ile Pro Phe Ala Gln Pro Pro Val Gly Asn   35

CTC CGC TTC AAG CCG CCT GTG CCG TAC TCG GCG TCT CTC AAT GGT CAG AAG TTT ACT CTG TAT GGC CCT CTG TGC   225
Leu Arg Phe Lys Pro Pro Val Pro Tyr Ser Ala Ser Leu Asn Gly Gln Lys Phe Thr Ser Tyr Gly Pro Ser Cys   60

ATG CAG ATG AAC CCA TTG GGC AAC TGG GAC TCC TCG CTT CCC AAG GCT GCC ATC AAC CTG TTG ATG CAG TCC AAG   300
Met Gln Met Asn Pro Leu Gly Asn Trp Asp Ser Ser Leu Pro Lys Ala Ala Ile Asn Ser Leu Met Gln Ser Lys   85

CTC TTC CAG GCG GTG CTT CCT AAC GGC GAG GAC TGT CTC ACC ATC AAC GTG GTG CGG CCG CTG GGC ACC AAG CCG   375
Leu Phe Gln Ala Val Leu Pro Asn Gly Glu Asp Cys Leu Thr Ile Asn Val Val Arg Pro Ser Gly Thr Lys Pro   110
```

```
GGT GCC AAC CTC CCC GTG ATG GTG TGG ATT TTT GGC GGC GGG TTT GAG GTT GGC GGC TCC AGT CTC TTC CCT CCC  450
Gly Ala Asn Leu Pro Val Met Val Trp Ile Phe Gly Gly Gly Phe Glu Val Gly Gly Ser Ser Leu Phe Pro Pro  135

GCA CAG ATG ATC ACC GCC AGC GTG CTT ATG GGC AAG CCC ATC ATC CAC GTG AGC ATG AAC TAC CGC GTT GCT TCG  525
Ala Gln Met Ile Thr Ala Ser Val Leu Met Gly Lys Pro Ile Ile His Val Ser Met Asn Tyr Arg Val Ala Ser  160

TGG GGG TTC TTG GCT GGT CCA GAC ATC AAG GCC GAG GGC AGC GGG AAC GCC GGT TTG CAC GAC CAA CGC TTG GGT  600
Trp Gly Phe Leu Ala Gly Pro Asp Ile Lys Ala Glu Gly Ser Gly Asn Ala Gly Leu His Asp Gln Arg Leu Gly  185

TTG CAG TGG GTG GCG GAC AAC ATT GCC GGG TTC GGC GGC GAC CCG TCC AAG GTG ACC ATC TTT GGT GAG CTG GCG  675
Leu Gln Trp Val Ala Asp Asn Ile Ala Gly Phe Gly Gly Asp Pro Ser Lys Val Thr Ile Phe Gly Glu Ser Ala  210
                                                                                  *   *   *   *

GGC AGC ATG TCG GTA ATG TGT CAG CTC CTC TGG AAC GAC GGC GAC AAC ACG TAC AAC GGC AAG CCG TTG TTC CGT  750
Gly Ser Met Ser Val Met Cys Gln Leu Leu Trp Asn Asp Gly Asp Asn Thr Tyr Asn Gly Lys Pro Leu Phe Arg  235
*

GCC GCC ATC ATG CAG CTG GGG GCC ATG GTG CCG CTG GAC CCG GTG GAT GGG CCC TAC GGC ACG CAG ATC TAC GAC  825
Ala Ala Ile Met Gln Ser Gly Ala Met Val Pro Ser Asp Pro Val Asp Gly Pro Tyr Gly Thr Gln Ile Tyr Asp  260

CAG GTG GTT GCT TCA GCC GGC TGT GGC AGT GCC AGC GAC AAG CTC GCG TGC TTG CGC AGC ATC CTG AAC GAC AAA  900
Gln Val Val Ala Ser Ala Gly Cys Gly Ser Ala Ser Asp Lys Leu Ala Cys Leu Arg Ser Ile Ser Asn Asp Lys  285

CTC TTC CAG GCC ACC AGC GAC ACT CCG GGG GCC TTG GCG TAC CCC TCG TTG CGG TTG CTG TTT CTC CCG CGG CCC  975
Leu Phe Gln Ala Thr Ser Asp Thr Pro Gly Ala Leu Ala Tyr Pro Ser Leu Arg Leu Ser Phe Leu Pro Arg Pro  310

GAC GGC ACC TTC ATC ACC GAT GAC ATG TTC AAG TTG GTG CGC GAC GGC AAG TGT GCC AAC GTT CCG GTG ATC ATT 1050
Asp Gly Thr Phe Ile Thr Asp Asp Met Phe Lys Leu Val Arg Asp Gly Lys Cys Ala Asn Val Pro Val Ile Ile  335

GGC GAC CAG AAC GAC GAG GGC ACA GTG TTT GCG TTG CTG CTG TTG AAC GTG ACT ACG GAT GCT CAG GCA CGC CAG 1125
Gly Asp Gln Asn Asp Glu Gly Thr Val Phe Ala Leu Ser Ser Leu Asn Val Thr Thr Asp Ala Gln Ala Arg Gln  360

TAC TTC AAG GAA CTG TTC ATC CAC GCC AGC GAC GCG GAG ATC GAC ACC TTG ATG GCG GCG TAC CCC AGC GAC ATC 1200
Tyr Phe Lys Glu Ser Phe Ile His Ala Ser Asp Ala Glu Ile Asp Thr Leu Met Ala Ala Tyr Pro Ser Asp Ile  385

ACC CAG GGT CTG CCG TTC GAC ACC GGC ATC TTC AAC GCC ATC ACC CCG CAG TTC AAA CGG ATT GCA GCG GTG CTT 1275
Thr Gln Gly Ser Pro Phe Asp Thr Gly Ile Phe Asn Ala Ile Thr Pro Gln Phe Lys Arg Ile Ala Ala Val Leu  410

GGT GAC CTT GCG TTC ACT CTC CCC CGG CGC TAC TTC CTC AAC CAC TTC CAG GGC GGC ACC AAG TAC TCG TTC CTC 1350
Gly Asp Leu Ala Phe Thr Leu Pro Arg Arg Tyr Phe Leu Asn His Phe Gln Gly Gly Thr Lys Tyr Ser Phe Leu  435

CTG AAG CAG CTT CTG GGG TTG CCG GTG ATT GGC ACC CAC CAC GCC AAC GAC ATT GTG TGG CAG GAC TTT TTG GTG 1425
Ser Lys Gln Leu Ser Gly Leu Pro Val Ile Gly Thr His His Ala Asn Asp Ile Val Trp Gln Asp Phe Leu Val  460

AGC CAC AGC AGC GCC GTG TAC AAC AAC GCG TTT ATT GCC TTT GCC AAC GAC CTC GAC CCG AAC AAG GCC GGT TTG 1500
Ser His Ser Ser Ala Val Tyr Asn Asn Ala Phe Ile Ala Phe Ala Asn Asp Leu Asp Pro Asn Lys Ala Gly Leu  485

CTT GTG AAC TGG CCC AAG TAC ACC AGC AGC CTG CAG CTG GGC AAC AAC TTG TTG CAG ATC AAC GCC TTG GGC TTG 1575
Leu Val Asn Trp Pro Lys Tyr Thr Ser Ser Ser Gln Ser Gly Asn Asn Leu Leu Gln Ile Asn Ala Leu Gly Leu  510

TAC ACC GGC AAG GAC AAC TTC CGC ACC GCT GGC TAC GAC GCG TTG TTT ACC AAC CCG CTG CTG TTT TTT GTT TAG 1650
Tyr Thr Gly Lys Asp Asn Phe Arg Thr Ala Gly Tyr Asp Ala Leu Phe Thr Asn Pro Ser Ser Phe Phe Val ●●● 534

GAACTTACAAGCGCCATACCGCTTCAGCCGTCTTGGC
```

C

```
                                                                                AAGGCACCAAGGCGGT
CAAGACACAAAGAGCGTGGGGAGGTGACAATGCCGATCGCGGACCAGTTTACACCGCTCCGAGTATAAAAGCAGAAGCATTCTCACCTGCTCGCTCCCC

ATG AAG CTC GCT CTT GCG CTC CTG CTC ATT GCC TCG GTG GCT GCT GCC CCC ACC GCC ACG CTC GCC AAC GGC GAC   75
Met Lys Leu Ala Leu Ala Leu Ser Leu Ile Ala Ser Val Ala Ala Ala Pro Thr Ala Thr Leu Ala Asn Gly Asp   10

ACC ATC ACC GGT CTC AAC GCC ATC ATC AAC GAG GCG TTC CTC GGC ATT CCC TTT GCC GAG CCG CCG GTG GGC AAC  150
Thr Ile Thr Gly Leu Asn Ala Ile Ile Asn Glu Ala Phe Leu Gly Ile Pro Phe Ala Glu Pro Pro Val Gly Asn   35

CTC CGC TTC AAG GAC CCT GTG CCG TAC CGT GGG TCT CTC AAC GGT CAA TCC TTC ACC GCG TAC GGT CCG CTG TGC  225
Leu Arg Phe Lys Asp Pro Val Pro Tyr Arg Gly Ser Leu Asn Gly Gln Ser Phe Thr Ala Tyr Gly Pro Ser Cys   60

ATG CAG CAG AAC CCC GAG GGC ACC TAC GAG GAG AAC CTC CCC AAG GTG GCG CTT GAC TTG GTG ATG CAG TCC AAG  300
Met Gln Gln Asn Pro Glu Gly Thr Tyr Glu Glu Asn Leu Pro Lys Val Ala Leu Asp Leu Val Met Gln Ser Lys   85

GTG TTC CAG GCT GTT CTC CCC AAC AGC GAG GAC TGC CTC ACC ATC AAC GTG GTG CGG CCG CCG GGC ACC AAG GCG  375
Val Phe Gln Ala Val Leu Pro Asn Ser Glu Asp Cys Leu Thr Ile Asn Val Val Arg Pro Pro Gly Thr Lys Ala  110

GGC GCC AAC CTC CCG GTC ATG CTC TGG ATC TTT GGC GGT GGG TTT GAG ATC GGC AGC CCC ACC ATC TTC CCT CCC  450
Gly Ala Asn Leu Pro Val Met Leu Trp Ile Phe Gly Gly Gly Phe Glu Ile Gly Ser Pro Thr Ile Phe Pro Pro  135

GCT CAG ATG GTC TCC AAG AGT GTG CTC ATG GGC AAG CCC ATC ATC CAC GTG GCC GTC AAC TAC CGC TTG GCG TCC  525
Ala Gln Met Val Ser Lys Ser Val Leu Met Gly Lys Pro Ile Ile His Val Ala Val Asn Tyr Arg Leu Ala Ser  160

TTT GGT TTC TTG GCC GGT CCG GAC ATC AAG GCC GAG GGC AGC TCC AAT GCC GGC CTC AAG GAC CAG CGC TTG GGC  600
Phe Gly Phe Leu Ala Gly Pro Asp Ile Lys Ala Glu Gly Ser Ser Asn Ala Gly Leu Lys Asp Gln Arg Leu Gly  185

ATG CAG TGG GTG GCA GAC AAC ATT GCC GGG TTC GGC GGC GAC CCG AGC AAG GTG ACC ATC TTT GGT GAG CTG GCG  675
Met Gln Trp Val Ala Asp Asn Ile Ala Gly Phe Gly Gly Asp Pro Ser Lys Val Thr Ile Phe Gly Glu Ser Ala  210
                                                                                  *   *   *   *

GGC AGC ATG TCC GTG TTG TGC CAC CTT CTC TGG AAT GGC GGC GAC AAC ACG TAC AAG GGC AAG CCG TTG TTC CGC  750
Gly Ser Met Ser Val Leu Cys His Leu Leu Trp Asn Gly Gly Asp Asn Thr Tyr Lys Gly Lys Pro Leu Phe Arg  235
*

GCG GGC ATC ATG CAG CTG GGA GCC ATG GTG CCG CTG GAC CCG GTG GAC GGC ACC TAT GGA ACC CAA ATC TAT GAC  825
Ala Gly Ile Met Gln Ser Gly Ala Met Val Pro Ser Asp Pro Val Asp Gly Thr Tyr Gly Thr Gln Ile Tyr Asp  260
```

RNSDOCID: EXP 10943018

```
ACG TTG GTG GCT CTG ACG GGC TGC AGC AGT GCC AGC AAC AAG CTT GCG TGC TTG CGT GGT CTT CTG ACT CAG GCA   900
Thr Leu Val Ala Ser Thr Gly Cys Ser Ser Ala Ser Asn Lys Leu Ala Cys Leu Arg Gly Leu Ser Thr Gln Ala   285

TTG CTC GAT GCC ACC AAC GAC ACC CCT GGG TTC TTG TCG TAC ACC TCG TTG CGG TTG CTG TAC CTT CCC CGG CCC   975
Leu Leu Asp Ala Thr Asn Asp Thr Pro Gly Phe Leu Ser Tyr Thr Ser Leu Arg Leu Ser Tyr Leu Pro Arg Pro   310

GAC GGC GCC AAC ATC ACC GAT GAC ATG TAC AAG TTG GTA CGC GAC GGC AAG TAT GCA AGC GTT CCC GTG ATC ATT  1050
Asp Gly Ala Asn Ile Thr Asp Asp Met Tyr Lys Leu Val Arg Asp Gly Lys Tyr Ala Ser Val Pro Val Ile Ile   335

GGC GAC CAG AAC GAC GAG GGC TTC TTG TTT GGT CTC CTG CTG TTG AAC ACC ACC ACC GAG GCC GAC GCC GAG GCA  1125
Gly Asp Gln Asn Asp Glu Gly Phe Leu Phe Gly Leu Ser Ser Leu Asn Thr Thr Thr Glu Ala Asp Ala Glu Ala   360

TAC CTC AGA AAG CTG TTC ATC CAC GCC ACC GAC GCC GAT ATC ACC GCA TTG AAG GCG GCG TAC CCC AGC GAT GTC  1200
Tyr Leu Arg Lys Ser Phe Ile His Ala Thr Asp Ala Asp Ile Thr Ala Leu Lys Ala Ala Tyr Pro Ser Asp Val   385

ACC CAG GGT CTG CCG TTC GAC ACG GGC ATT CTC AAC GCC CTT ACA CCC CAG CTC AAG CGG ATC AAT GCT GTG CTT  1275
Thr Gln Gly Ser Pro Phe Asp Thr Gly Ile Leu Asn Ala Leu Thr Pro Gln Leu Lys Arg Ile Asn Ala Val Leu   410

GGC GAC CTC ACC TTT ACC CTC TCG CGC CGC TAC TTC CTC AAC CAC TAC ACC GGT GGT CCC AAG TAC TCG TTC CTC  1350
Gly Asp Leu Thr Phe Thr Leu Ser Arg Arg Tyr Phe Leu Asn His Tyr Thr Gly Gly Pro Lys Tyr Ser Phe Leu   435

CTG AAG CAG CTT CTG GGC TTG CCC ATT CTC GGC ACG TTC CAC GCG AAC GAC ATT GTG TGG CAG CAC TTT TTG TTG  1425
Ser Lys Gln Leu Ser Gly Leu Pro Ile Leu Gly Thr Phe His Ala Asn Asp Ile Val Trp Gln His Phe Leu Leu   460

GGC AGC GGC AGC GTC ATC TAC AAC AAC GCG TTC ATC GCG TTT GCC ACC GAC TTG GAC CCC AAC ACC GCG GGC TTG  1500
Gly Ser Gly Ser Val Ile Tyr Asn Asn Ala Phe Ile Ala Phe Ala Thr Asp Leu Asp Pro Asn Thr Ala Gly Leu   485

CTG GTG CAG TGG CCC AAG TCC ACC AGC AGC CTG CAG GCG GGG GAC AAC TTG ATG CAG ATC AGT GCC TTG GGC TTG  1575
Ser Val Gln Trp Pro Lys Ser Thr Ser Ser Ser Gln Ala Gly Asp Asn Leu Met Gln Ile Ser Ala Leu Gly Leu   510

TAC ACC GGC AAG GAC AAC TTC CGC ACC GCC GGC TAC AAC GCT TTG TTT GCC GAC CCG CTG CAC TTT TTC GTG TAG  1650
Tyr Thr Gly Lys Asp Asn Phe Arg Thr Ala Gly Tyr Asn Ala Leu Phe Ala Asp Pro Ser His Phe Phe Val ••• 534

ACAATGTACATGTGCACCCCCAAGAAGCCTGCTTCGTCCACCAGCCC
```

Fig. 2. The nt sequences of *LIP3* (**A**), *LIP4* (**B**) and *LIP5* (**C**) and deduced aa sequences. The nt are numbered starting from the start codon as position A = +1. The aa are numbered taking the first aa of the mature proteins as position +1 (to the right of black arrowhead). TATAA and CAAT-like sequences are underlined. The black arrowhead indicates predicted signal peptidase cleavage sites. Lip consensus sequences are marked by asterisks, and potential *N*-glycosylation sites are boxed. CTG-Ser are in bold-face type. The nt sequences were determined by the dideoxy chain-termination method (Sanger et al., 1977). Sequences have been submitted to the EMBL Databank under accession Nos. X66006 (*LIP3*), X66007 (*LIP4*) and X66008 (*LIP5*).

and CAAT boxes enclosed in conserved sequence regions (Fig. 3). CAAT and TATAA boxes are implicated in the initiation of transcription in eukaryotes and have been found upstream from genes highly expressed in *Candida* species (Singer et al., 1989). Further evidence of Lip expression was inferred from the comparison with the restriction maps of Ccl cDNAs isolated by other authors (Kawaguchi and Honda, 1991).

The predicted pIs (4.5 for Lip1, 4.9 for Lip2, 5.1 for Lip3, 5.7 for Lip4 and 5.5 for Lip5) would support a

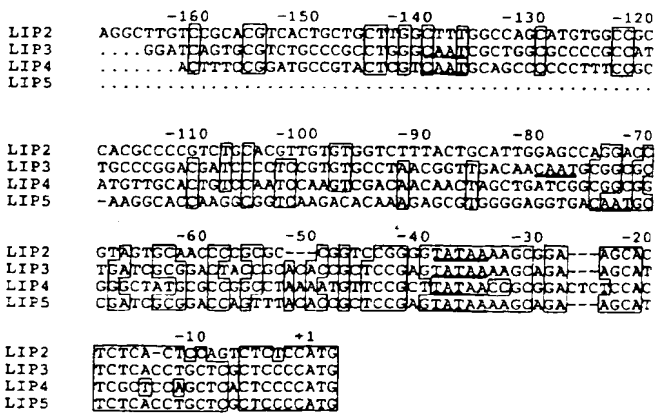


Fig. 3. Sequences upstream from *LIP2, 3, 4,* and *5*. TATAA and CAAT boxes are underlined. The nt conserved in at least three sequences are boxed. Last digits of numerals are aligned with corresponding nt.

correlation among the putative proteins encoded by the genes isolated in our laboratory and the Lip isoforms detected in the enzyme purified preparations by different experimental approaches, such as two-dimensional gel electrophoresis (Veeraragavan et al., 1989) and affinity chromatography (E. Cernia, G. Ortaggi, M. Castagnola and R. Rubino, personal communication). Although glycosylation or other post-translational modifications could affect the pI predicted for the protein moiety to some extent, it is likely that the information provided by the sequence analysis could be useful for the separation and identification of *C. cylindracea* Lip isoforms.

In yeast, genes belonging to multigene families have been reported that are expressed under different growth conditions or produce enzymes differing in their substrate preference (Pretorius et al., 1991). Interestingly, a strain of *G. candidum* synthesizes Lip with a different substrate specificity (Sidebottom et al., 1991). The production of Lip isozymes could therefore be of physiological relevance in conferring to the yeast versatility in adaptation to different environments.

(d) Multiple alignment with homologous sequences

Screening of data bases by the FASTA program showed that the Lip from *C. cylindracea* are very weakly related to most other Lip, with the only exception being

BNSDOCID: <XP 1084201A >

the two Lip isoforms from *G. candidum* (40–44% identity over 544 aa aligned). Moreover, significant homology was observed to proteins belonging to the ChE family. The highest sequence similarities were found with AChE from *T. californica* (TAChE) and *T. marmorata* (27–31% identity over 446 aa aligned). The three-dimensional structures of both Gcl and TAChE have been recently solved by x-ray diffraction (Schrag et al., 1991; Sussman et al., 1991).

In order to gain some structural information concerning Ccl, sequence analysis of the Lip has been performed. This analysis can take advantage of two conditions: (*i*) multiple sequences belonging to the same gene family are available, (*ii*) structures of two different proteins, both homologous to Ccl, have been described. Fig. 4 shows the multiple alignments of the aa sequences encoded by *LIP* (Lip1–5), the sequences reported for two *G. candidum* isoforms (Shimada et al., 1989; 1990) and best scoring esterase sequences in the FASTA output. Sequences of the mature proteins have been considered.

(e) Secondary structure elements

Sequence homology, when available, is at present the most reliable tool for secondary structure prediction. The identity relating the Ccls to Gcls and TAChE is greatly above the threshold of significance (25% identity for more than 80 residues aligned) introduced by Sander and Schneider (Sander and Schneider, 1991). In this case, the reliability of the method employed for the alignment can be assessed, since two sequences of known structure have been enclosed. As shown in Fig. 4, corresponding domains in the secondary structures of Gcls and TAChE are indeed aligned. The good superimposition of the two reference sequences with those of Lip suggests that the Ccl might have a similar structural organization.

(f) The helical lid

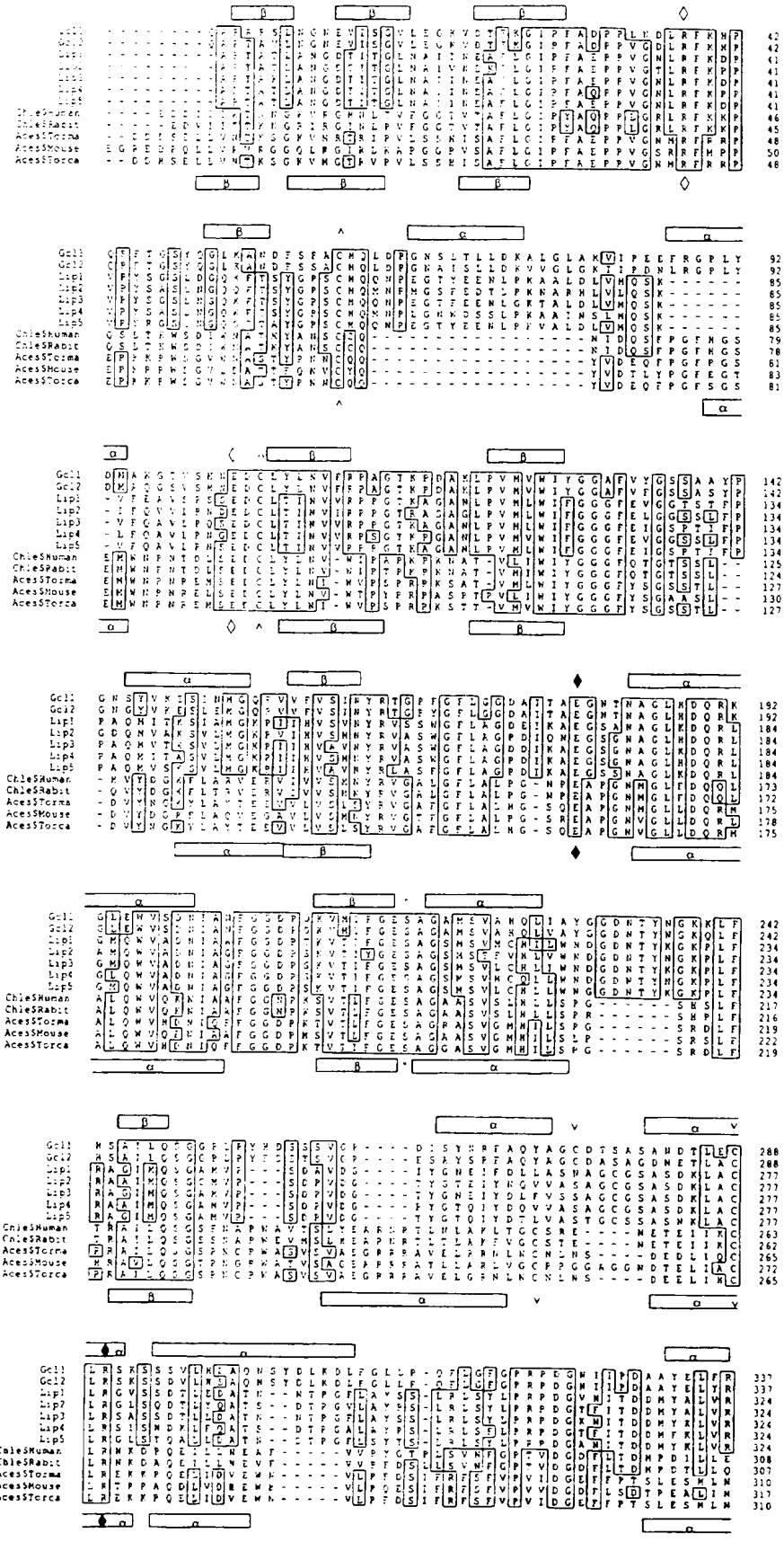
Despite the high overall similarity, local sequence divergences are observed in Lip in comparison with the Gcl. In particular, the sequences of the first, second and eighth helices of Gcl cannot be satisfactorily aligned to the corresponding regions of the Lip sequences, even if gaps are introduced (Fig. 4). The lack of significant homology in these regions suggests that they might have a different structural organization in Lip. It should be noted, however, that several statistical methods for secondary structure prediction all consistently predict a helical conformation for the region 63–94 of the Lip sequences (data not shown). These helices have been described to build up the 'lid' covering the active site of the Gcl and proposed to undergo conformational changes during the process of interfacial activation of this enzyme (Schrag et al., 1991). It is intriguing that the high and continuous sequence homology between the two fungal Lip is abruptly interrupted in these regions. It could be hypothesized that such local divergences form the basis of the different substrate specificity of the two enzymes. Corresponding aa stretches are not conserved in the AChE sequences. In these cases, the lack of a structure analogous to the lid has been demonstrated by crystallography. Based only on the alignment of Fig. 4 and in the absence of direct structural information, it is not possible to hypothesize whether other structural elements might provide the 'lid' function to Ccl. It is also possible that the regions corresponding to the 'lid' elements can protect the active site or that conformational changes different from those hypothesized for the other Lip have an analogous functional role for the mechanism of interfacial activation in *Candida* Lip.

(g) Structural organization

Key residues stabilizing the folding of Gcl and TAChE are conserved in Ccl (Fig. 4). Charged residues forming salt bridges in Gcl (Arg^{38}/Glu^{103} and Glu^{180}/Arg^{290}) and in TAChE (Arg^{44}/Glu^{92} and Glu^{163}/Arg^{267}) are conserved at equivalent positions in the Ccl family. This would suggest a role for the corresponding pairs Arg^{37}/Glu^{95} and Glu^{172}/Arg^{279} in the stabilization of the tertiary structure of the Ccl (Fig. 4). Cys residues forming two disulfide bridges in Gcl (Cys^{61}/Cys^{105} and Cys^{276}/Cys^{288}) and two out of the three disulfide bridges in TAChE (Cys^{67}/Cys^{104} and Cys^{255}/Cys^{265}) are conserved also in the Lip family. The only exception concerns Cys^{268} of Ccl, Cys^{254} of TAChE and Cys^{276} of Gcl that are shifted by a few residues with respect to each other and to their partners in the disulfide bridge (Cys^{277}, Cys^{265} and Cys^{288}, respectively). It is therefore likely that the frame of disulfide bridges is maintained in the three enzymes, where in Ccl the four Cys we hypothesize to be connected are Cys^{60}/Cys^{97} and Cys^{268}/Cys^{277} (Fig. 4). The third disulfide bridge of AChE is not present in fungal Lip. The strong conservation of residues involved in protein structure stabilization is suggestive of a general similarity of the three-dimensional structures of the three enzymes.

(h) The catalytic triad

Crystallographic data on the Gcls and TAChE pointed out the presence of a trypsin-like catalytic triad in their active site. Residues composing the catalytic triad of Gcl (Ser^{217}, Glu^{354} and His^{463}; Schrag et al. 1991) and of TAChE (Ser^{200}, Glu^{327} and His^{400}; Sussmann et al., 1991) are conserved in the Lip sequences (Fig. 4). Moreover, these residues are surrounded by highly conserved regions. Our results strongly support the hypothesis



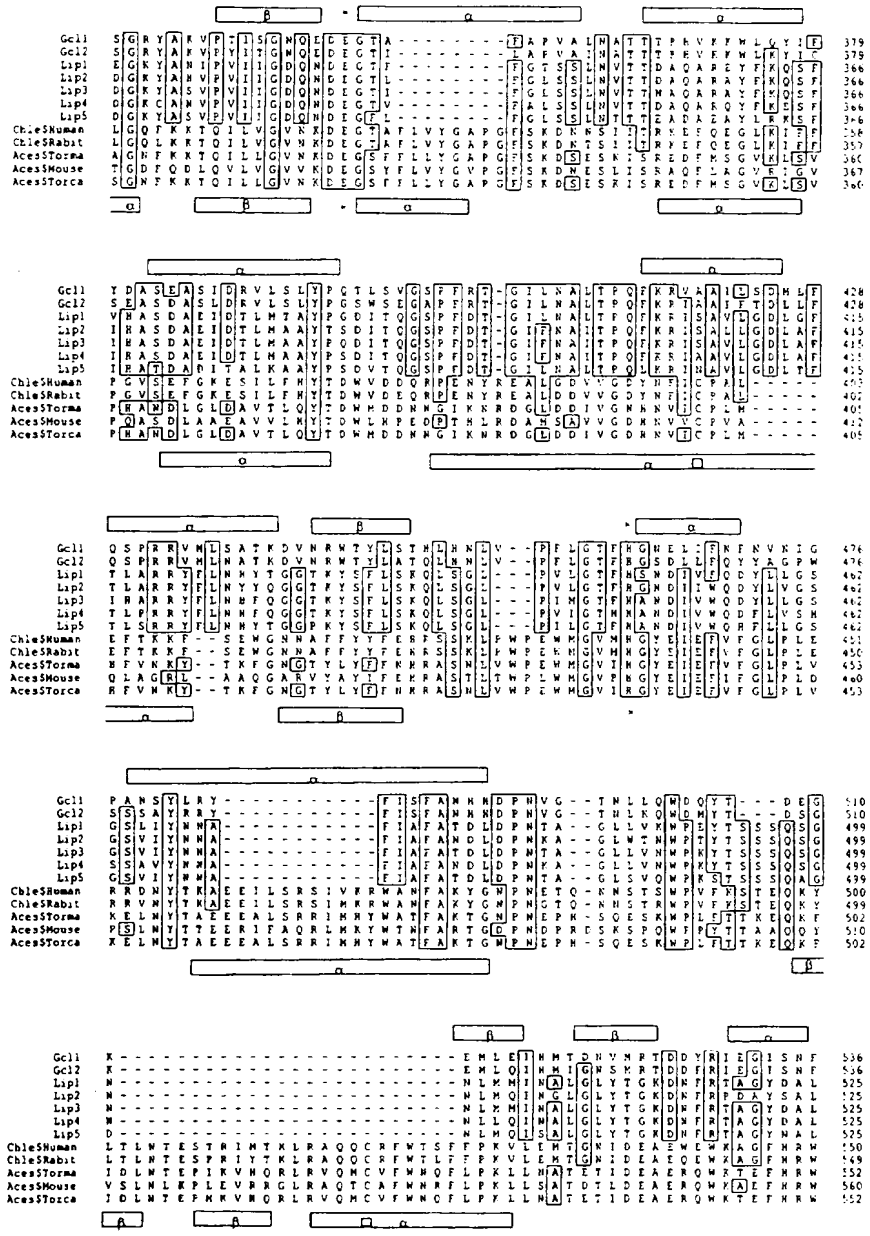


Fig. 4. Multiple alignment (Devereux et al., 1984) of the putative Lip (Lip1–5), two Lip isoforms from *G. candidum* (Gcl1 and Gcl2), ChE from *Homo sapiens* (ChleSHuman) and *Rattus norvegicus* (Chle$Rabit) and AChE from *T. marmorata* (Aces$Torma), mouse (Aces$Mouse) and *T. californica* (Aces$Torca). The aa are numbered starting from the first aa of the mature proteins. Identical aa conserved among more than 50% of the sequences are boxed. Symbols on the top and the bottom of each sequence block refer to Gcl1 and Aces$Torca, respectively. Active site residues are indicated by asterisks, secondary structure elements by open bars, residues involved in salt bridges by diamonds, where the same filling pattern identifies partner residues in the bridge. Paired Cys are indicated by upward-pointing carets, downward-pointing carets, and open squares (the last is present only in AChE).

(Schrag et al., 1991) that Ser^{209}, Glu^{341} and His^{449} are part of the active site of Ccl.

(i) Evolutionary analysis

The active Ser residue of the superfamily of serine hydrolases is embedded in the consensus motif $GlyXaa_1Ser-Xaa_2Gly$, where in Lip Xaa_1 is usually either Tyr or His. The consensus sequence within the AChE family is reported to be PheGlyGluSerAlaGly (Bairoch, 1990). Interestingly, the catalytic Ser of the two fungal Lip is surrounded by a consensus sequence characteristic of AChE. This motif is highly conserved in the Lip sequences, the only exception being Lip2 in which a Tyr substitutes for Phe. This observation – if substantiated by the isolation of an active product of *LIP2* – would suggest either that in the first position of the consensus sequence any aromatic side chain might be allowed or it might correlate with differences in catalytic function or in substrate specificity. The carboxylic chain of the catalytic triad appears to be provided by glutamate instead of aspartate, as in other Ser hydrolases. AChE (Sussman et al., 1991), the Gcl (Schrag et al., 1991) and the Ccl are the only examples reported to date of Ser hydrolases using a Ser-His-Glu catalytic triad. Recently, Krejci and collegues (1991) proposed grouping in the same protein family proteins containing a common ChE-like domain. This family comprises several Ser hydrolases, including esterases, AChE and Gcl. On the basis of the sequences determined for *LIP*, Ccl would also belong to this family.

These observations, taken together with the overall sequence homology among Gcls, Ccls and AChE, suggest a common ancestor gene for these enzymes. The two fungal Lip could therefore be considered as a bridge between the esterases and the Lip family (Fig. 5).

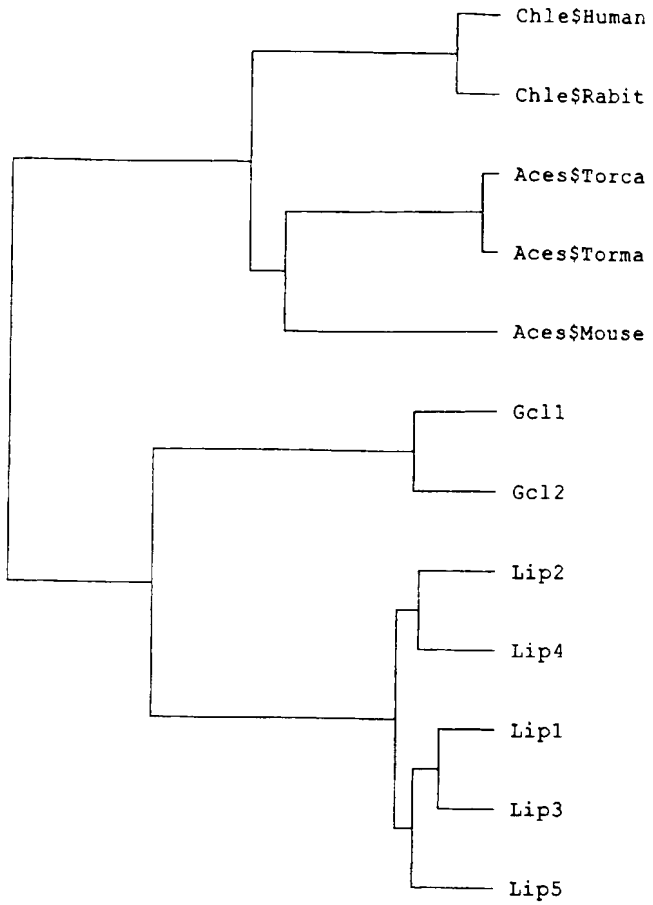


Fig. 5. Tree representation (Devereux et al., 1984) of the clustering relationships among the sequences aligned (and explained) in Fig. 4.

(j) The genetic code

A unique feature of the *C. cylindracea* strain ATCC14830 from which we have cloned *LIP* is the use of a non-universal genetic code. By comparing nt and aa sequences, it has been demonstrated that in this strain the codon CTG codes for Ser instead of Leu (Kawaguchi et al., 1989). Interestingly, CTG is used with a high frequency (3% of the codons) in *LIP*, including that corresponding to the catalytic Ser. Out of the 24 CTG-Ser, 13 are conserved in the Lip family. At non-conserved positions, CTG-Ser were found to be substituted by Ser (other codons), Leu, Ala, Pro, Trp, Asn or His. How such a phenomenon could originate and be maintained throughout evolution remains to be clarified.

(k) Conclusions

(*1*) In the yeast *C. cylindracea* are present several Lip genomic sequences (at least five but possibly even more).

(*2*) The availability of multiple genes belonging to the same family provides an useful tool for the identification of structurally significant sequence features. The more interesting observation emerging from the comparison of the Lip sequences is the conservation of the consensus sequence Phe/Tyr-Gly-Glu-Ser-Ala-Gly surrounding the active Ser^{209}, supporting a close relation of the *C. cylindracea* enzymes with the ChE family.

(*3*) By the use of sequence alignments we have inferred a tentative model of the structural organization of Ccl based mainly on the crystallographic data reported for the homologous Gcl. The overall structures of the two fungal Lip are likely to be very similar, except for the helical chains building up the 'lid' in Gcl.

ACKNOWLEDGEMENTS

This work has been partially supported by a grant of the CNR Progetto Finalizzato Biotecnologie e Biostru-

3NSDOCID <XP 1084201A >

mentazioni and by the BRIDGE Program of the Commission of the European Communities.

REFERENCES

Alberghina, L., Grandori, R., Longhi, S., Lotti, M., Fusetti, F. and Vanoni, M.: Molecular cloning of a Lip and of a Lip-related gene from *Candida cylindracea*. In: Alberghina, L., Schmid, R. and Verger, R. (Eds), Lipases: Structure, Mechanism and Genetic Engineering, GBF Monographs No. 16. VCH, Weinheim, 1991, pp. 231–235.

Antonian, E.: Recent advances in the purification, characterization and structure determination of Lips. Lipids 23 (1988) 1101–1106.

Bairoch, A.: PROSITE: A dictionary of protein sites and patterns. University of Geneva, Geneva, 1990.

Boel, E., Huge-Jensen, B., Christensen, M., Thim, L. and Fiil, N.P.: *Rhizomucor miehei* triglyceride Lip is synthesized as a precursor. Lipids 23 (1988) 701–706.

Brady, L., Brzozowski, A.M., Derewenda, Z.S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J.P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thim, L. and Menge, U.: A serine protease triad forms the catalytic centre of a triacylglycerol Lip. Nature 343 (1990) 767–770.

Brockman, H.L.: General features of lipolysis: reaction scheme, interfacial structure and experimental approaches. In: Borgstrom, B. and Brockman, H.L. (Eds) Lipases. Elsevier, NY, 1984, pp. 3–46.

Brzozowski, A.M., Derewenda, U., Derewenda, Z.S., Dodson, G.G., Lawson, D.M., Turkenburg, J.P., Bjorkling, F., Huge-Jensen, B., Patkar, S.A. and Thim, L.: A model for interfacial activation in Lips from the structure of a fungal Lip-inhibitor complex. Nature 351 (1991) 491–494.

Chapus, C., Sémériva, M., Bovier-Lapierre, C. and Desnuelle, P.: Mechanism of pancreatic Lip action, I. Interfacial activation of pancreatic Lip. Biochemistry 15 (1976) 4980–4987.

Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 (1984) 387–395.

Harwood, J.: The versatility of Lips for industrial uses. Trends Biochem. Sci. 14 (1989) 125–126.

Hubbard, S.C. and Ivatt, R.J.: Synthesis and processing of asparagine-linked oligosaccharides. Annu. Rev. Biochem. 50 (1981) 555–583.

Kawaguchi, Y. and Honda, H.: The structure of Lip genes and pseudogenes of *Candida cylindracea*. In: Alberghina, L., Schmid, R. and Verger, R. (Eds), Lipases: Structure, Mechanism and Genetic Engineering, GBF Monographs No. 16. VCH, Weinheim, 1991, pp. 221–230.

Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S.: The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*. Nature 341 (1989) 164–166.

Krejci, E., Duval, N., Chatonnet, A., Vincens, P. and Massoulié, J.: Cholinesterase-like domains in enzymes and structural proteins: functional and evolutionary relationships and identification of a catalytically essential aspartic acid. Proc. Natl. Acad. Sci. USA 88 (1991) 6647–6651.

Longhi, S., Fusetti, F., Grandori, R., Lotti, M., Vanoni, M. and Alberghina, L.: Cloning and nucleotide sequences of two Lip genes from *Candida cylindracea*. Biochim. Biophys. Acta 1131 (1992) 227–232.

Pretorius, I.S., Lambrechts, M.G. and Marmur J.: The glucoamylase multigene family in *Saccharomyces cerevisiae* var. *diastaticus*: an overview. Crit. Rev. Biochem. Mol. Biol. 26 (1991) 53–76.

Sander, C. and Schneider, R.: Database of homology-derived protein structures and structural meaning of sequence alignment. Proteins 9 (1991) 56–68.

Sanger, F., Niklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Schrag, J.D., Li, Y., Wu, S. and Cygler, M.: Ser-His-Glu triad forms the catalytic site of the Lip from *Geotrichum candidum*. Nature 351 (1991) 761–764.

Shimada, Y., Sugihara, A., Tominaga, Y., Iizumi, T. and Tsunasawa, S.: cDNA molecular cloning of *Geotrichum candidum* Lip. J. Biochem. 106 (1989) 383–388.

Shimada, Y., Sugihara, A., Iizumi, T. and Tominaga, Y.: cDNA cloning and characterization of *Geotrichum candidum* Lip II.J. Biochem. 107 (1990) 703–707.

Sidebottom, C.M., Charton, E., Dunn, P.P.J., Mycock, G, Davies, C., Sutton, J.L., Macrae, A.R. and Slabas, A.R.: *Geotrichum candidum* produces several Lips with markedly different substrate specificities. Eur. J. Biochem. 202 (1991) 485–491.

Singer, S.C., Richards, C.A., Ferone, R., Benedict, D. and Ray, P.: Cloning, purification, and properties of *Candida albicans* thymidylate synthase. J. Bacteriol. 171 (1989) 1372–1378.

Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. and Silman, I.: Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. Science 253 (1991) 872–879.

Tomizuka, N., Ota, Y. and Yamada, K.: Studies on Lip from *Candida cylindracea*, Part II. Amino acid composition, carbohydrate component and some physical properties. Agric. Biol. Chem. 30 (1966) 1090–1096.

Veeraragavan, K. and Gibbs, B.F.: Detection and partial purification of two Lips from *Candida rugosa*. Biotechnol. Lett. 11 (1989) 345–348.

von Heijne, G.: A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14 (1986) 4683–4690.

Winkler, F.K., D'Arcy, A. and Hunziker, W.: Structure of human pancreatic Lip. Nature 343 (1990) 771–774.